Econometrica, Vol. 64, No. 1 (January, 1996), 51–75

MULTIPRODUCT NONLINEAR PRICING

BY MARK ARMSTRONG[1]

Typically, work on mechanism design has assumed that all private information can be captured in a single scalar variable. This paper explores one way in which this assumption can be relaxed in the context of the multiproduct nonlinear pricing problem. It is shown that the firm will choose to exclude some low value consumers from all markets. A class of cases that allow explicit solution is derived by making use of a multivariate form of "integration by parts." In such cases the optimal tariff is cost-based.

KEYWORDS: Nonlinear pricing, mechanism design.

1. INTRODUCTION

THE PAST TWO DECADES have seen a rapid advance in the theory of mechanism design, for instance in the areas of nonlinear pricing, monopoly regulation, taxation, and the design of auctions. With some notable exceptions this work has modelled informational asymmetries by assuming that all private information can be captured in a single scalar variable. This paper explores one way in which this assumption may be relaxed in the context of the multiproduct nonlinear pricing problem.

The main results of previous work on the single-product nonlinear pricing problem have been (i) discovering ways to solve for the optimal tariff; (ii) showing that in many cases the firm will wish to separate completely its customers, so that customers with different tastes buy different quantities; (iii) showing that those customers with the strongest preferences for the good are served efficiently, with others being served lesser quantities than would be efficient in a world with full information, and (iv) showing that in many cases it is optimal for the firm to offer quantity discounts, so that the marginal price for a unit of the good decreases with the total quantity purchased.

An interesting question is how far any of these results from the single-product nonlinear pricing problem extend to multiproduct monopoly, or, more generally, what exactly will a profit-maximizing multiproduct nonlinear tariff look like? Questions it would be desirable to answer include: When is the optimal multiproduct tariff concave? (This would be one generalization of the single-product quantity discount result.) When does the optimal tariff involve the marginal price of one good decreasing with consumption of the other goods? (This would be another kind of generalization of quantity discounts.) And what does the pattern of demand look like? For instance, to what extent does a

[1] Thanks to Grant Hillier, Paul Klemperer, Jim Mirrlees, Patrick Rey, Kevin Roberts, Jean-Charles Rochet, Bob Wilson, and to Rob Porter and three referees for many helpful comments. Financial assistance from the ESRC (Grant Number L114251026) is gratefully acknowledged.

consumer's decision to purchase one good affect her decisions over other goods? However, before answering any of these we must first find a method of solving for the optimal multiproduct nonlinear tariff, at least within a class of cases.

The starting points for the present paper are Section 4 of Mirrlees (1976) and Section 7 of Mirrlees (1986). In the context of optimal tax theory, he sets out the framework of continuously distributed multidimensional types used here, and derives some first-order conditions for an incentive scheme to be optimal. However, no examples are solved, nor are any hints given on what the optimal tax regime might look like in such a setting.

Papers in the area of multiproduct price discrimination are few. Most often the framework is simplified to that of "unit demands," where consumers buy either one unit of a given good or none at all.[2] Moving beyond the assumption of unit demands, a paper that has both multiple products and multiple dimensions of types is Spence (1980). Indeed, consumer i in his model simply has a utility function $u_i(x)$, and so consumers are not parameterized at all (other than by i). He models consumers discretely and solves only a subproblem: Given that the various consumers are each going to be assigned a certain bundle, what is the best way to make them pay for these bundles? He has, however, no results concerning the optimal method of assigning these bundles in the first place. Nevertheless, he does highlight a central problem in solving the multidimensional case: provided the "single-crossing" property holds, in the single-good/single-parameter case it is possible to give a natural ordering of consumers since higher types choose higher quantities and, under a wide variety of conditions, only the "downward" incentive compatibility constraints will bind. In Spence's model it is not possible to predict in advance which of the many incentive constraints will bind, and this makes the problem rather intractable.

In contrast to Spence's paper, Roberts (1979) and Mirman and Sibley (1980) analyze the multiproduct nonlinear pricing problem under the assumption that consumers differ only by a scalar parameter (although in their conclusion Mirman and Sibley pose the problem of multidimensional types). Such a problem may be solved using methods similar to the standard single-product case, but I would argue that such models shed little light on how to design multiproduct tariffs. Since consumers differ only by a scalar, the chosen bundles (given any tariff) will all lie on some one-dimensional curve in the product space, and so it will not be possible to address the question of to what extent consumers are encouraged or discouraged to buy two products rather than one. Moreover, it is rarely the case that knowledge of a given consumer's taste for one product determines exactly what her tastes are for any other (although this is not to deny that tastes for different goods could well be correlated).

Laffont, Maskin, and Rochet (1985) consider the opposite problem to that of Roberts and Mirman and Sibley: the firm offers only a single product but

[2]This case of "bundling" is discussed by Adams and Yellen (1976), Schmalensee (1984), and McAfee et al. (1989), amongst others. See also Example 1 below.

consumers are differentiated by a two-dimensional parameter. As a result there will necessarily be "bunching," and some consumers with different tastes for the good will buy the same quantity. It is interesting to see that in their example the optimal tariff has quantity discounts and the consumer with the strongest taste for the good is efficiently served. These properties, then, are not purely the result of an assumption that consumers differ only by a scalar parameter. The analysis required to solve their example is *much* more complicated than that needed for the scalar problem, and this indicates that mathematical difficulties occur when it is the type space rather than the decision space which is multidimensional. Again, though, their model is not capable of answering the question of what makes a good multiproduct tariff.

McAfee and McMillan (1988) is perhaps the paper that in technical terms is closest to this one. There are three distinct sections, and the first proposes a "generalized single-crossing property" which guarantees that necessary conditions for a differentiable incentive scheme to be implementable are also sufficient when agents differ by a multidimensional parameter. Second, they use this generalized single-crossing property to extend the paper by Laffont et al. to a more general framework and they solve the single-good/multiple-type problem for a class of cases (the earlier paper simply found an example). In their third section they analyze the problem of a multiproduct monopolist selling indivisible goods to a single buyer. They show that in the case of two goods and under a kind of bivariate hazard-rate condition it is not necessary to use a stochastic selling procedure, and that three prices (one for each good purchased in isolation and one for the bundle of two goods) is all that is needed to attain the optimum. More relevant for this paper is that they implicitly use the procedure of "integrating along rays" that I follow in Section 4.2.

Wilson (1991; 1993, Chapters 12–14; and 1995) analyzes a similar model to that presented here. He derives first-order conditions for an incentive scheme to be optimal, although like Mirrlees he has no general results concerning the shape of the optimal tariff nor of the pattern of demand it induces. Wilson (1993 and 1995) also uses numerical simulations in order to obtain particular solutions to the problem. Finally, Wilson (1993) has a chapter on real-world examples of multiproduct nonlinear tariffs, which include advertising rates in U.S. periodicals, electricity tariffs in France, telephone tariffs in America, and the tariffs offered by express mail delivery firms.

The plan of this paper is as follows. In Section 2 I set out a model for analyzing multiproduct nonlinear pricing. In Section 3 it is shown that the firm will usually choose to exclude a nontrivial set of low-demand consumers from the market, a result that contrasts with the scalar case. In Section 4 I set out a strategy for solving the multiproduct, multidimensional problem which follows closely the usual scalar approach. As with the scalar case, this involves ignoring the "implementability" constraint on the consumer demand function, and then checking afterwards that this constraint has been satisfied. In the scalar case this constraint was that consumer demands had to be increasing in type, whereas in the multidimensional case matters are much more complex. In all cases for which this procedure is valid the optimal charge for a bundle of goods is simply a function of the total cost of producing the bundle.

Finally, we have to make a judgment about whether to model consumers discretely or continuously. The former has the advantage of being more intuitive and "realistic," but notationally is messier, for instance in its use of inequalities to describe the various incentive compatibility constraints. The paper by Spence takes this approach. If, on the other hand, we choose to have a continuous distribution of types then, firstly, we will be able to use calculus to more powerful effect, and secondly there will then be a continuum of different bundles chosen—this means we will get a tariff *schedule* instead merely of an assignment of charges to a finite number of bundles. This will probably be more illuminating and I therefore choose the latter.

2. A MODEL

Consider a firm with a monopoly over n goods. Consumers of these goods have a variety of preferences over these goods parameterized by the m-dimensional vector $\alpha = (\alpha_1, \ldots, \alpha_m)$. Suppose that a type-$\alpha$ consumer's utility if she consumes the bundle of goods $x = (x_1, \ldots, x_n)$ and makes a payment t is

$$u(\alpha, x) - t \tag{1}$$

where $u(0, x) = u(\alpha, 0) = 0$, u is increasing in all arguments, and u is continuous, convex, and homogeneous of degree one in α.[3]

Some justification for these severe restrictions is clearly needed. The main reason to choose this form of utility is that it is probably the simplest utility function that involves nontrivial multidimensional types. At present there is little understanding of multidimensional mechanism design problems and it seems a good strategy to start simply and build more complex models when our intuition is better developed. (For instance, early accounts of single-product nonlinear pricing used the utility function (1) with $n = m = 1$—see Mussa and Rosen (1978).) Secondly, a class of utility functions that satisfy all these conditions are those that take the form

$$u(\alpha, x) = \sum_{i=1}^{n} \alpha_i u_i(x_i). \tag{2}$$

Therefore, in these cases the dimension of the type space is equal to the number of products, utility is linear in α, there are no cross-price effects in demand, and there is one parameter α_i associated in a natural way with each product i.[4] Such

[3] All of the following analysis can simply be modified to allow u to be homogeneous of arbitrary degree $d > 0$ in α, but this adds little of interest.

[4] For instance, I believe that it is easier to use the utility function (2) than to have multidimensional types but only a *single* product, the case analyzed by Laffont et al. (1985) and McAfee and McMillan (1988). The reason for this is that the type parameters can be treated in a more symmetric manner, and this allows for more straightforward solutions.

utility functions are the natural extensions of consumer utility in the "bundling" model developed by Adams and Yellen which has been more thoroughly analyzed, and which also has these restrictions.

Assume that the firm incurs a cost $c(x)$ in serving any consumer with the bundle of products x, where $c(0) = 0$ and c is smooth. Therefore, the firm's total cost function is separable across consumers. A particular case of such a cost function is when the firm has constant marginal cost c_i for producing a unit of product i, in which case $c(x) = \sum_{i=1}^{n} c_i x_i$. There may be economies of scope in serving a given customer with several products, however, and so we will use the more general formulation. In order to guarantee that any consumer's consumption choice is well-defined in the following analysis, we assume that the *efficient* allocation to the type-α consumer is finite, i.e. that a solution to the problem of maximizing $u(\alpha, x) - c(x)$ over $x \geq 0$ exists for all α.

The firm is not able to observe a given consumer's type, but has prior beliefs over the distribution of types, described by the density function $f(\alpha)$ which has support $A \subset R_{+}^{m}$. For technical reasons A is closed, convex, and has full dimension in R^m, and f is continuously differentiable on A. The firm aims to maximize its profit by offering a tariff $t(\cdot)$, where $t(x)$ is its charge for the bundle of quantities x.[5] Suppose that there are no possibilities for arbitrage among consumers and that the firm is able to monitor the sales to each consumer. Therefore the firm may offer tariffs that take any form (including those that involve quantity discounts or quantity premia). If consumers cannot be forced to consume, the firm is required to set $t(0) \leq 0$.

Faced with a particular tariff t, a consumer of type α obtains a surplus of

(3) $$s(\alpha) \equiv \max_{x \geq 0} : u(\alpha, x) - t(x).$$

Since $t(0) \leq 0$ this function $s(\cdot)$ is nonnegative. It is also necessarily continuous, increasing, and (because of the convexity of u in α) convex in α. If $s(\cdot)$ is differentiable at α it will satisfy the envelope condition

(4) $$\nabla s(\alpha) = u_\alpha(\alpha, x(\alpha)),$$

where $\nabla s(\alpha) \equiv (\partial s(\alpha)/\partial \alpha_1, \ldots, \partial s(\alpha)/\partial \alpha_m)$ is the vector derivative of s evaluated at α, $x(\alpha)$ is the (necessarily unique) choice of quantities that maximizes utility for the type-α consumer given the tariff t, and u_α denotes the vector partial derivative of u with respect to α. The fact that s is convex means that it is differentiable almost everywhere (in the sense of Lebesgue measure) on A. Since consumers are continuously distributed on A, this implies there will be a unique optimal choice of quantities for almost all consumers. Therefore, we will use the notation $x(\alpha)$ for the consumer demand function even when this function is not necessarily everywhere well defined.

[5] Unlike McAfee and McMillan (1988) I do not consider the possibility of the firm using stochastic selling strategies.

Not all functions $x(\alpha)$ can be the result of optimizing behavior by consumers when faced with a nonlinear tariff. We say that a demand function which can be induced by a suitable tariff is *implementable*:

DEFINITION: *The surplus function* $s(\cdot)$ *and demand function* $x(\cdot)$ *are said to be implementable by the firm if they may be induced from* (3) *and* (4) *by means of some tariff* $t(\cdot)$ *such that* $t(0) \leq 0$.

As is well-known, in the scalar, single-product case where the "single-crossing" condition holds, a necessary and sufficient condition for a demand function to be implementable is that it be increasing in the type parameter. Unlike McAfee and McMillan (1988) and Rochet (1987), this paper is not concerned with finding necessary and sufficient conditions for implementability in the multidimensional case, and when we find optimal demand functions in Section 4 we will ensure that they are indeed implementable simply by constructing a tariff which induces this behavior.[6]

Since it is almost always the case that $t(x(\alpha)) = u(\alpha, x(\alpha)) - s(\alpha)$, the profit obtained by the firm from a type-α consumer is $u(\alpha, x(\alpha)) - s(\alpha) - c(x(\alpha))$, where $x(\alpha)$ and $s(\alpha)$ are related by (4). The total profit is then the sum of all of these individual profit contributions:

$$\pi = \int_A [u(\alpha, x(\alpha)) - s(\alpha) - c(x(\alpha))] f(\alpha)\, d\alpha. \tag{5}$$

Therefore, the central problem we shall be concerned with is to maximize (5) subject to (4) and s and x being implementable. Thus we work with the induced surplus function s rather than the tariff t, a procedure usually adopted in similar adverse selection models following the work of Mirrlees (1971).

Before we turn to a class of cases that admit an explicit solution in Section 4, we describe one result that holds fairly generally.

3. THE DESIRABILITY OF EXCLUSION

PROPOSITION 1: *If* $m \geq 2$ *and* A *is strictly convex, then at the optimum a set of consumers of positive measure will not buy any goods.*

PROOF: See Appendix A.

The intuition for this result is straightforward. If a tariff is such that all consumers *do* participate in the market, consider what happens to profits if a small fixed charge ε is added to this tariff. The firm will gain ε from every consumer who remains in the market, but will pay the penalty of causing some

[6] However, when utility takes the simple, separable form in (2) above, it is fairly straightforward to show that the necessary and sufficient conditions for implementability are (i) that the surplus function $s(\alpha)$ is convex, and (ii) that $\partial u_i(x_i)/\partial \alpha_j \equiv \partial u_j(x_j)/\partial \alpha_i$.

low demand consumers to exit. In the multidimensional case where A is strictly convex, for small ε the number who exit is (usually) of order strictly greater than ε (i.e. it is of order ε^2, say).[7] Therefore, the former effect will dominate the latter, and a degree of exclusion will prove optimal.[8] (In the single-product case the number who exit typically is of order ε, and so there is no way that we can deduce that this latter effect is small in relation to the former.) Even if the demand parameters are very high in relation to cost, the firm will not wish to serve all of the market. In effect, for any tariff which (just) induces participation from all consumers an increase in the tariff will not cause many consumers to exit—i.e. demand (or more precisely, participation) is inelastic at that point with respect to lump-sum increases in the tariff—and standard microeconomics tells us that it cannot be optimal for a monopolist to charge a price that results in inelastic demand.[9]

Proposition 1 does *not* hold generally in the scalar context, and so the phenomenon of exclusion provides a major difference between scalar and multidimensional nonlinear pricing. For instance, an examination of (12) below shows that the single-product firm will choose to serve all consumers if the distribution of consumer tastes is high enough in relation to costs. This difference is also illustrated in the following simple example:

EXAMPLE 1: *The bundling model with two goods: consumers' valuations are uniformly distributed on the unit square* $[a, a+1] \times [a, a+1]$, *and the firm has costless production.*

Consumers have unit demands for two goods, and a consumer of type (α_1, α_2) gains a (gross) utility from consuming a single unit of good i of α_i (and gains no further utility from further consumption of that good).[10] Then without loss of generality the firm sets three prices: p_i is the charge for consuming a unit of good i in isolation, and p_b is the charge for consuming both goods. Since consumers' valuations are symmetrically distributed we examine symmetric

[7] This informal explanation of the result ignores one technical caveat which is dealt with in the formal proof. At the optimum it must be that at least one type of consumer receives no surplus (for otherwise the firm could increase the tariff by some small amount and not cause exit). If almost all consumers do buy some goods, but the marginal consumer, for some reason, is served with zero quantities (i.e. if the surplus function s is flat when $s = 0$), then it may be that by increasing the tariff by ε will cause a number of consumers of order ε to exit. If this knife-edge situation occurs, then the intuition given in the text is not valid. However, the proof demonstrates that if this occurs, then the profit lost by the exit of these consumers is of order strictly greater than ε, which is all that is needed for the argument to work.

[8] This result depends crucially on the assumption of a continuous distribution, and if consumers were modeled discretely this result would not necessarily hold. (For instance, if all consumers were of the same type, then the firm typically would wish to serve the entire market.)

[9] We may deduce that at the optimum, the "elasticity of participation" with respect to lump-sum increases in the tariff must be greater than one.

[10] Formally, consumers have utility function given by (2) and $u_i(x_i) = 0$ if $0 \le x_i < 1$ and $u_i(x_i) = 1$ if $x_i \ge 1$.

tariffs where $p_1 = p_2 = p$, say. There are then two broad alternative policies: the firm can choose a subadditive tariff, where $p_b \le 2p$ and consumers are given an incentive to buy both goods, or the firm can offer a superadditive tariff, where $p_b > 2p$. The analysis of McAfee et al. (1989) suggests that when consumer valuations are independently distributed (as they are here) the firm will choose to offer a subadditive tariff, in which case the pattern of demand may appear as in Figure 1. Here we assume that $a \le p \le a+1$ and $p_b \ge p + a$. Region R_i in the figure denotes those consumer types who buy only good i, R_b is the region where both goods are purchased, and R_0 is the exclusion region. In this case, by calculating the areas of the various regions, the firm's total profit if it offers such a tariff is

$$\pi = 2p(a+1-p)(p_b - p - a) + p_b[(a+1-p)(a+1+p-p_b) + \tfrac{1}{2}(2p-p_b)(2a+2-p_b)].$$

(The first term in the above is the profit from those consumers who buy just one good and the second term is the profit from those who buy both.)

On the other hand, when $p \ge p_b - a$ the regions R_1 and R_2 vanish (i.e., there is pure bundling), profit depends only on p_b, and in the case where $2a \le p_b \le 2a + 1$ profit is given by

$$\pi = p_b\left[1 - \tfrac{1}{2}(p_b - 2a)^2\right].$$

Profit is a piecewise cubic in (p, p_b) which makes it a nonconcave problem, but when $a > 0$ it can be calculated that a (local) maximum of this profit function is obtained by setting

$$p_b = \tfrac{1}{3}\left(4a + \sqrt{4a^2 + 6}\right); \qquad p \ge p_b - a.$$

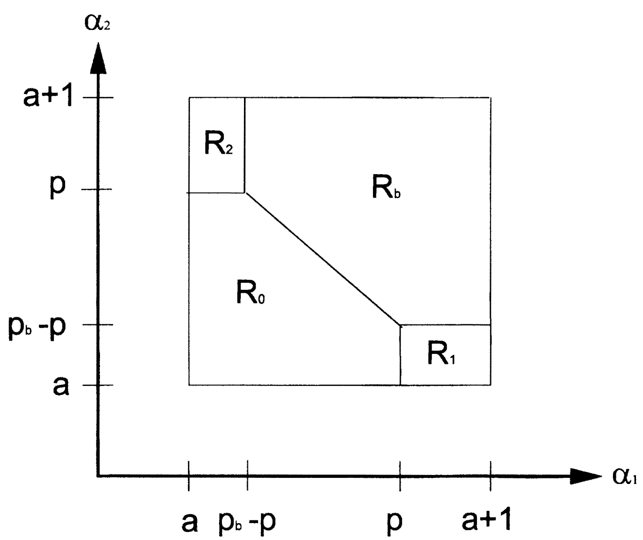

FIGURE 1.—A subadditive tariff without pure bundling.

(See Figure 2). Thus there is pure bundling and the fraction of consumers who are excluded is

$$\tfrac{1}{18}\left(\sqrt{4a^2+6}-2a\right)^2.$$

This fraction tends to zero as a becomes large, but it is always optimal to exclude some (relatively) low value consumers no matter how far away the support of consumer valuations lies from the origin. This contrasts with the single-good version of this example, in which it can easily be seen that the firm would choose to serve all consumers if $a \geq 1$.

4. MULTIDIMENSIONAL NONLINEAR PRICING

4.1. *Nonlinear Pricing: The Scalar Case*

The method by which some multiproduct solutions can be obtained follows closely the standard method used in the single-product, single-characteristic case, and so it is worth recapitulating briefly that technique. (For more details see, for instance, Mussa and Rosen (1978), Maskin and Riley (1984), and Wilson (1993).) In this case the utility function is just $\alpha u(x)$, where both α and x are scalars. Profits are still given by (5) except that here A is an interval $[\alpha_*, \alpha^*]$, and so

$$\pi = \int_{\alpha_*}^{\alpha^*} [\alpha u(x(\alpha)) - s(\alpha) - c(x(\alpha))] f(\alpha)\, d\alpha. \tag{6}$$

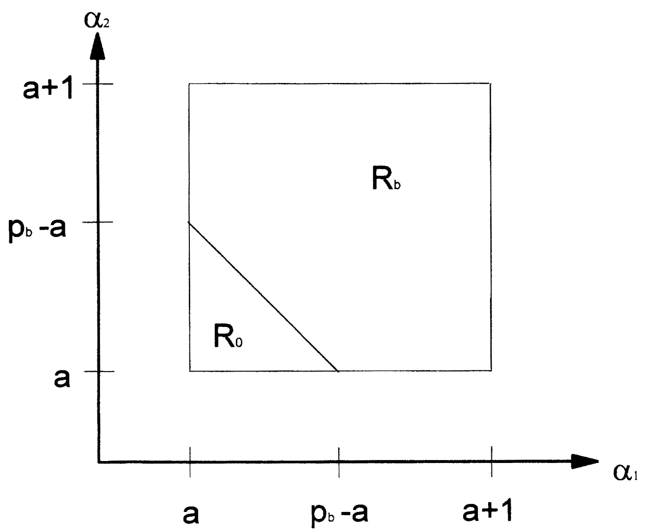

FIGURE 2.—The optimal pattern of demand.

The problem is to maximize this expression subject to the constraints that s is nonnegative, that the envelope condition $u(x(\alpha)) = s'(\alpha)$ holds, and that $x(\cdot)$ is nondecreasing. (This last constraint ensures that the demand function is implementable by some tariff.)

Since $u(x(\alpha)) = s'(\alpha)$ we can use integration by parts to obtain

$$(7)\qquad \int_{\alpha_*}^{\alpha^*} s(\alpha) f(\alpha)\, d\alpha = \int_{\alpha_*}^{\alpha^*} u(x(\alpha))[1 - F(\alpha)]\, d\alpha$$

where $F(\cdot)$ is the distribution function corresponding to the density $f(\cdot)$ and we have used the fact that $s(\alpha_*) = 0$ (as must be optimal since profit in (6) is decreasing in $s(\cdot)$). Therefore we can express (6) as

$$(8)\qquad \pi = \int_{\alpha_*}^{\alpha^*} \{[\alpha u(x(\alpha)) - c(x(\alpha))] f(\alpha) - u(x(\alpha))[1 - F(\alpha)]\}\, d\alpha$$

which gives profit in terms of the demand function $x(\cdot)$. All that needs to be done now is to maximize this subject to the constraint that $x(\cdot)$ be nonnegative and nondecreasing. (Since we have built into this the assumption that $s(\alpha_*) = 0$ and $s' = u(x) \geq 0$, it follows automatically that $s(\cdot) \geq 0$.) If we ignore the constraint that $x(\cdot)$ be nondecreasing we can simply maximize the integrand in (8) pointwise to obtain the following candidate for the optimal demand function x^*:

$$(9)\qquad x^*(\alpha) \underset{x \geq 0}{\text{maximizes}}: \left[\alpha - \frac{1 - F(\alpha)}{f(\alpha)}\right] u(x) - c(x).$$

It is then simple to show that $x^*(\cdot)$ in (9) is nondecreasing—and hence that all constraints are satisfied and that this procedure is valid—if the following condition on the distribution on the scalar variable α holds:[11]

$$(10)\qquad \alpha - \frac{1 - F(\alpha)}{f(\alpha)} \text{ is nondecreasing.}$$

To see when quantity discounts (in the sense that the optimal tariff t is concave) are optimal we argue as follows. The first-order condition for the type-α consumer to demand quantity $x^*(\alpha)$ is that $\alpha u'(x^*(\alpha)) = t'(x^*(\alpha))$, and from the first-order condition for $x^*(\alpha)$ to solve (9) we obtain

$$t'(x^*(\alpha)) = \frac{c'(x^*(\alpha))}{1 - (1 - F(\alpha))/(\alpha f(\alpha))}.$$

[11] If this hazard rate condition does not hold then the "ironing" procedure described in Mussa and Rosen (1978) will be required and bunching will prove optimal.

Therefore, provided that $c(x)$ is weakly concave and that the stronger condition

$$(11) \qquad 1-\frac{1-F(\alpha)}{\alpha f(\alpha)} \text{ is nondecreasing}$$

holds, then (i) condition (10) holds and so $x^*(\cdot)$ is nondecreasing, and (ii) $t'(x^*(\cdot))$ is nonincreasing in α. Because $x^*(\cdot)$ is nondecreasing this implies that $t'(x)$ is nonincreasing in x, i.e. that $t(\cdot)$ is concave.[12]

Finally, in order to justify the claim made in Section 3 that exclusion was not always optimal in the single-product setting, we can see from (9) that the firm will choose to serve *all* consumers provided that

$$(12) \qquad \left[\alpha_* - \frac{1}{f(\alpha_*)}\right] u'(0) > c'(0)$$

(where α_* is the lowest taste parameter). Therefore, shifting the support of consumer tastes far enough up the real line must necessarily cause the firm to serve all consumers, in contrast to the multidimensional case.

4.2. *Nonlinear Pricing: The Case of Multiple Characteristics*

Here I mirror the technique described above. For this it is convenient to allow α to be distributed over the entire nonnegative orthant R^m_+, although the density $f(\alpha)$ may vanish on portions of this set. Just as with the single-dimensional case, we wish to be able to use some kind of "integration by parts" in order to express the expected value of $s(\alpha)$ in terms of the demand function $x(\alpha)$. One way to do this is the following. Since it must be optimal to set $s(0)=0$, we have

$$s(\alpha)=\int_0^1 \frac{d}{dr} s(r\alpha)\, dr.$$

The envelope condition (4) implies that

$$\frac{d}{dr} s(r\alpha) = \alpha' u_\alpha(r\alpha, x(r\alpha)) = \frac{1}{r} u(r\alpha, x(r\alpha)),$$

where the final equality follows from the homogeneity of u in α. (This is the reason we assume u to have this property.) We therefore obtain

$$(13) \qquad s(\alpha)=\int_0^1 \frac{1}{r} u(r\alpha, x(r\alpha))\, dr.$$

[12]A sufficient condition for both the above conditions on the distribution of α to hold is that the *hazard rate* of α, which is $f(\alpha)/[1-F(\alpha)]$, be nondecreasing. In turn, a sufficient condition for α to have a nondecreasing hazard rate is that its density function $f(\alpha)$ be log-concave—for details see Fudenberg and Tirole (1991, Chapter 7).

Using (13) we obtain the following:

$$\int_{R_+^m} s(\alpha)f(\alpha)\,d\alpha = \int_{R_+^m}\left[\int_0^1 \frac{1}{r}u(r\alpha, x(r\alpha))\,dr\right]f(\alpha)\,d\alpha$$
$$= \int_0^1\left[\int_{R_+^m} \frac{1}{r}u(r\alpha, x(r\alpha))f(\alpha)\,d\alpha\right]dr.$$

Making the change of variables $\hat{\alpha} = r\alpha$ and using the fact that $d\hat{\alpha} = r^m\,d\alpha$ gives

$$\int_{R_+^m} \frac{1}{r}u(r\alpha, x(r\alpha))f(\alpha)\,d\alpha = \int_{R_+^m} \frac{1}{r^{m+1}}u(\hat{\alpha}, x(\hat{\alpha}))f(\hat{\alpha}/r)\,d\hat{\alpha}$$

and so

$$\int_{R_+^m} s(\alpha)f(\alpha)\,d\alpha = \int_0^1\left[\int_{R_+^m} \frac{1}{r^{m+1}}u(\hat{\alpha}, x(\hat{\alpha}))f(\hat{\alpha}/r)\,d\hat{\alpha}\right]dr$$
$$= \int_{R_+^m} u(\hat{\alpha}, x(\hat{\alpha}))\left[\int_0^1 \frac{f(\hat{\alpha}/r)}{r^{m+1}}\,dr\right]d\hat{\alpha}.$$

By making the substitution $t = 1/r$ we obtain

$$\int_0^1 \frac{f(\alpha/r)}{r^{m+1}}\,dr = \int_1^\infty t^{m-1}f(t\alpha)\,dt.$$

In sum:

$$\int_{R_+^m} s(\alpha)f(\alpha)\,d\alpha = \int_{R_+^m} u(\alpha, x(\alpha))g(\alpha)\,d\alpha \tag{14}$$

where

$$g(\alpha) = \int_1^\infty t^{m-1}f(t\alpha)\,dt. \tag{15}$$

Expression (14) is the multidimensional version of (7) above.[13]

Using (14) we can express the firm's expected profit (5) as

$$\pi = \int_{R_+^m} \{[u(\alpha, x(\alpha))' - c(x(\alpha))]f(\alpha) - u(\alpha, x(\alpha))g(\alpha)\}\,d\alpha, \tag{16}$$

which, as with the single-good case, expresses profit in terms of the demand function $x(\cdot)$. Therefore, maximizing the above pointwise with respect to x leads

[13] It is straightforward to show that when $m = 1$, $g(\alpha) = [1 - F(\alpha)]/\alpha$. Also, I assume throughout the remainder of this paper that enough moments for α exist for the function $g(\alpha)$ always to be finite.

to the following candidate for the optimal demand function $x^*(\alpha)$:

$$(17)\qquad x^*(\alpha)\ \underset{x\geq 0}{\text{maximizes}}: \left[1-\frac{g(\alpha)}{f(\alpha)}\right]u(\alpha, x)-c(x).$$

This corresponds to (9) above. Thus, just as with the scalar case, this candidate demand function results in consumers being served with less than efficient levels of outputs except on the upper boundary of the support of α when they are served efficient levels (for in such cases $g(\alpha)=0$).

As in the scalar case, this procedure ignores the constraint that the demand function x^* must be able to be implemented by means of a suitable tariff. In the scalar case this involved the relatively straightforward requirement that x^* be nondecreasing in type. In the multiple characteristics case, however, matters are much more complex. If we can verify that the demand function in (17) is implementable, though, all constraints have been satisfied and x^* is indeed the optimal allocation.

In order for these demands to be implementable we need to make two assumptions, both involving a kind of separability. The first is the following: Let $V(\alpha, y)$ be the "cost-based" indirect utility function associated with the utility function $u(\alpha, \cdot)$, i.e.

$$V(\alpha, y)\equiv \max: \{u(\alpha, x)|c(x)\leq y\}.$$

Then our assumption is that this function is separable in characteristics and income:

$$(18)\qquad V(\alpha, y)=h(\alpha)v(y)$$

where $h(\alpha)$ is a function that is necessarily homogeneous of degree one given that u is. Examples that fit into this framework are discussed in Section 4.6 below.

Given (18), the maximization problem (17) may be expressed as the type-α consumer choosing outputs $x^*(\alpha)$ where

$$(19)\qquad x^*(\alpha)\ \text{maximizes: } u(\alpha, x)\ \text{subject to } c(x)\leq y^*(\alpha),$$

and

$$(20)\qquad y^*(\alpha)\ \underset{y\geq 0}{\text{maximizes}}: \left[1-\frac{g(\alpha)}{f(\alpha)}\right]h(\alpha)v(y)-y.$$

Before discussing further when the demands in (19) and (20) can be implemented it is useful to have a digression on optimal "cost-based" tariffs.

4.3. *Cost-based Tariffs*

Say that a multiproduct tariff $t(x)$ is *cost-based* if it takes the form $t=t(c(x))$, where $t(\cdot)$ is a function of a scalar variable. In other words, the firm makes the charge for consuming any bundle x depend only upon its cost of producing that bundle. Given such a tariff, a type-α consumer will choose quantities $x(\alpha)$ to

$$\underset{x \geq 0}{\text{maximize}}: u(\alpha, x) - t(c(x)),$$

which, given (18), implies that she will choose to purchase a bundle that costs $y(h(\alpha))$, where

$$y(h) \underset{y \geq 0}{\text{maximizes}}: hv(y) - t(y),$$

and to choose quantities $x(\alpha)$ that maximize $u(\alpha, x)$ subject to $c(x) \leq y(h(\alpha))$. As in the standard single-product case, the function $y(h)$ can be implemented by a cost-based tariff if and only if $y(\cdot)$ is nondecreasing in h.

Using the procedure outlined in Section 4.1 above, the firm's profit with the cost-based tariff $t(\cdot)$ is

$$\int_0^\infty \{[hv(y(h)) - y(h)]\phi(h) - v(y(h))[1 - \Phi(h)]\}\, dh,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the density and distribution functions for the random variable $h(\alpha)$ induced by the underlying density function $f(\alpha)$ for α. Therefore, the optimal function $y^*(h)$ is obtained by maximizing the above integrand pointwise provided that

$$\text{(21)} \qquad h - \frac{1 - \Phi(h)}{\phi(h)} \text{ is nondecreasing.}$$

Moreover, provided that

$$\text{(22)} \qquad 1 - \frac{1 - \Phi(h)}{h\phi(h)} \text{ is nondecreasing,}$$

then the optimal cost-based tariff $t(c(x))$ is concave in c. Finally, provided that $c(x)$ is a weakly concave function of x we deduce that if (22) holds then the optimal cost-based tariff is a concave function of x.

We summarize this discussion as a proposition:

PROPOSITION 2: *Given that the separability condition* (18) *and the hazard rate condition* (21) *hold, the optimal cost-based tariff results in the type-*α *consumer choosing outputs* x^* *where*

$$x^*(\alpha) \textit{ maximizes}: u(\alpha, x) \textit{ subject to } c(x) \leq y^*(h(\alpha)),$$

and

$$y^*(h) \underset{y \geq 0}{\textit{maximizes}}: \left[h - \frac{1 - \Phi(h)}{\phi(h)}\right] v(y) - y.$$

Moreover, if $c(x)$ is weakly concave and the stronger hazard rate condition (22) *holds, then the optimal cost-based tariff is concave in x.*

4.4. *A Soluble Class of Cases*

When, then, is $x^*(\alpha)$ as defined in (19) and (20) implementable? As well as requiring assumption (18) to hold, for the candidate demand function to be implementable we also need a second separability condition. Thus, suppose that f may be written in the multiplicatively separable form

$$f(\alpha)=f_1(h(\alpha))\times f_2(\alpha), \tag{23}$$

where h is defined in (18) and $f_2(\cdot)$ is homogeneous of degree zero in α. In words, this condition states that the fact a given consumer has a particular level of "average tastes" (as measured by the parameter $h(\alpha)$) gives us no further information about which of her taste parameters α_i are likely to be greater than others (i.e. on which ray from the origin α lies).[14]

Given (23), the density function for the parameter $h(\alpha)$ induced by $f(\alpha)$ can be shown to be

$$\phi(h)=kh^{m-1}f_1(h). \tag{24}$$

(See Appendix B for the derivation of this and the definition of k.) Therefore, the distribution function for h is just

$$\Phi(h)=k\int_0^h \hat{h}^{m-1}f_1(\hat{h})\,d\hat{h}. \tag{25}$$

Moreover, (23) implies that $g(\alpha)$ in (15) may be written

$$g(\alpha)=\left(\int_1^\infty t^{m-1}f_1(th(\alpha))\,dt\right)\times f_2(\alpha).$$

By changing variables, we may express this as

$$g(\alpha)=\left(\frac{1}{(h(\alpha))^m}\int_{h(\alpha)}^\infty \hat{h}^{m-1}f_1(\hat{h})\,d\hat{h}\right)\times f_2(\alpha). \tag{26}$$

In sum, the term hg/f which appears in (20) may be written as

$$\frac{h(\alpha)g(\alpha)}{f(\alpha)}=\frac{1-\Phi(h(\alpha))}{\phi(h(\alpha))}, \tag{27}$$

which is just the reciprocal of the hazard rate of the parameter h.

[14] For instance, consider the case of two goods and $h(\alpha)=\|\alpha\|$, where $\|\cdot\|$ is the standard Euclidean norm. Without loss of generality we can write f in terms of polar coordinates: $f=f(r,\theta)$, where $r=\|\alpha\|$ and θ is the angle the vector α makes with the horizontal axis. Then the above condition requires that f can be written as $f_1(r)f_2(\theta)$, i.e. that the two parameters r and θ which determine a consumer's tastes be independently distributed.

Therefore, given assumption (23) the candidate for the optimal demand function given by (19) and (20) may be expressed as the type-α consumer choosing outputs x^* where

$$x^*(\alpha) \text{ maximizes: } u(\alpha, x) \text{ subject to } c(x) \le y^*(h(\alpha)),$$

and

$$y^*(h) \underset{y \ge 0}{\text{maximizes}}: \left[h - \frac{1 - \Phi(h)}{\phi(h)}\right] v(y) - y.$$

But provided that (21) holds, this is precisely the demand function that corresponds to the optimal cost-based tariff described in Proposition 2 above. In particular, this demand function can be implemented by a suitable tariff (which is a cost-based tariff).

We summarize this discussion in the following proposition:

PROPOSITION 3: *Suppose that the two separability conditions* (18) *and* (23) *hold, and that the hazard rate condition* (21) *holds* (*where* ϕ *and* Φ *are defined in* (24) *and* (25) *respectively*). *Then the optimal demand function is as given by Proposition* 2 *above, and the optimal tariff is a cost-based tariff. In particular, if the stronger hazard rate condition* (22) *holds and if the cost function* $c(x)$ *is concave in* x, *then the optimal multiproduct tariff is concave in* x.

The difference between Propositions 2 and 3, of course, is that the former finds the optimal *cost-based* tariff (something that requires no special assumptions about the distribution of types other than the hazard rate condition (21)), whereas the latter demonstrates that whenever the additional separability condition (23) holds the optimal tariff over *all* tariffs is a cost-based tariff.

Given (23), sufficient conditions for the distribution of h to satisfy both hazard rate conditions (21) and (22) are easily found. Each of these conditions is satisfied provided that h has an increasing hazard rate. A sufficient condition for h to have an increasing hazard rate is that the density function for h, which from (24) is $kh^{m-1}f_1(h)$, be log-concave (see Section 4.1). In turn, a sufficient condition for this to be true is that $f_1(h)$ be log-concave.

There is a relatively simple explanation for the analysis used in Proposition 3. The method of integrating $s(\alpha)$ by parts along rays from the origin is precisely equivalent to the following procedure. Suppose that we cut the region R_+^m into many thin segments—see Figure 3. We could then treat each of these segments as a separate region and try to find the optimal demand function in each of the segments treated in isolation. Provided that the segments are sufficiently thin we can treat this problem as (approximately) a one-dimensional problem in h, the distance from the origin. (This was the problem analyzed by Mirman and Sibley (1980).) Integrating $s(\alpha)$ by parts along the ray in the familiar way we deduce

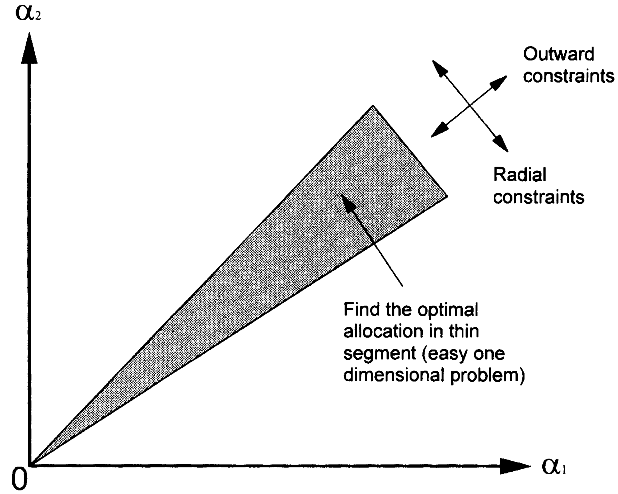

FIGURE 3.—Finding the optimum ignoring the "radial" incentive compatibility constraints.

that profit within the segment is analogous to (16):

$$\pi_{\text{segment}} = \int_{\text{segment}} \{[u(\alpha, x(\alpha)) - c(x(\alpha))]f(\alpha) - u(\alpha, x(\alpha))g(\alpha)\}\, d\alpha,$$

where g is given by (15). As before we can then just maximize the integrand pointwise with respect to x to obtain the candidate optimal demand function (17). Provided the hazard-rate condition which guarantees $s(\alpha)$ is convex along the ray holds, this demand function is definitely the optimal demand function for the firm *given that the support of types is just this thin segment*. (All the incentive compatibility conditions in the "outward" direction have been satisfied, and because of the thinness of the segment there are no "radial" incentive compatibility requirements.) The problem when we try to piece together these segments to obtain a solution for the whole orthant is that some consumers may be tempted to purchase bundles designated for those on neighboring segments. In other words, the "radial" incentive compatibility conditions have been ignored completely in this procedure. Under some circumstances these radial constraints will *not* bind and this procedure will then pick out the solution—a sufficient condition for this to hold is (23). In other cases radial incentive compatibility is an issue and the procedure of integrating along rays will not on its own pick out the solution.

4.5. *Possible Extensions to the Analysis*

Proposition 3 provides only a partial solution to the problem of finding the optimal multiproduct tariff, and we can speculate as to how further progress might be made when the density of types *cannot* be written in the separable form (23), or when the indirect utility function cannot be written in the separable form (18). There are at least four natural ways to try to extend the analysis.

First, we could use paths other than rays from the origin. The method for solving the single-product problem involved expressing the expected value of consumer surplus $s(\alpha)$ in terms of the demand function $x(\alpha)$ by integrating by parts. This involved writing $s(\alpha) = \int_0^\alpha u_\alpha(\hat{\alpha}, x(\hat{\alpha}))\, d\hat{\alpha}$. This section proposed one generalization of this procedure by expressing $s(\alpha)$ as the integral of the vector $u_\alpha(\alpha, x(\alpha))$ along the path given by the straight line joining α to the origin. However, there is no special reason for choosing this particular path, and $s(\alpha)$ could just as well be taken to be the integral of $u_\alpha(\alpha, x(\alpha))$ along *any* path joining α to the origin. Exactly the same method could be followed using another family of paths and this would generate a different "$g(\cdot)$" function to that given by (15). In principle, another condition analogous to (23), but covering another class of cases, could then be derived which would guarantee that the resulting candidate for the optimal demand function could be implemented by a tariff.

A second way forward might be to continue to use the "straight-line" path formula (16) for the firm's profits, but to introduce into the integral multipliers for the constraint that $x(\alpha)$ must be implementable. For instance, when utility takes the form (2), implementability requires that

$$\frac{\partial u_i(x_i)}{\partial \alpha_j} \equiv \frac{\partial u_j(x_j)}{\partial \alpha_i}. \tag{28}$$

In the two-good case, at least, this is straightforward to do, and it is possible to obtain first-order conditions for the optimal demands in terms of the multiplier for the single constraint (28).

On the other hand, when the separability condition (18) does not hold, it is no longer a straightforward matter even to find the optimal *cost-based* tariff as described in Section 4.3. In fact, in this case the problem of finding the optimal cost-based tariff is precisely the same problem of finding the optimal tariff for the single-product, multiple characteristics case which is analyzed in Laffont et al. (1985) and McAfee and McMillan (1988). In principle, then, it is possible at least within a class of cases to solve for the optimal cost-based tariff when (18) does not hold, and then to find conditions on the density function which guarantee that this tariff is the optimum over all tariffs.

Finally, a way to generate the optimal nonlinear tariff when neither (18) nor (23) holds is to use numerical techniques for particular specifications of costs and taste distributions. For an examination of the various possibilities involved here, see Wilson (1995).

4.6. *Solved Examples*

EXAMPLE 2: The multiproduct, scalar characteristic case. Suppose that $u(\alpha, x) = \alpha U(x)$, where α is a scalar and U is a utility function defined on the consumption of n products.[15] This satisfies (18) trivially since

$$V(\alpha, y) = \alpha v(y),$$

where

$$v(y) = \max: \{U(x) | c(x) \leq y\}.$$

In this case, Proposition 3 states that, provided the standard scalar hazard rate condition (10) holds for α, the optimal multiproduct tariff is a cost-based tariff.[16] Because α is a scalar, the second separability condition (23) is satisfied trivially.

EXAMPLE 3: Suppose now that $u(\alpha, x) = \sum_{i=1}^{n} \alpha_i x_i^{\gamma}$, where $0 < \gamma < 1$, and $c(x) = \sum_{i=1}^{n} c_i x_i$. This again satisfies (18) since

$$V(\alpha, y) = h(\alpha) y^{\gamma},$$

where

$$h(\alpha) = \left(\sum_{i=1}^{n} \frac{\alpha_i^{1/(1-\gamma)}}{c_i^{\gamma/(1-\gamma)}} \right)^{1-\gamma}.$$

Take the special case where $\gamma = \frac{1}{2}$ and $c_i \equiv c$. The second separability condition (23) is satisfied trivially whenever the density function f is a function only of $h(\alpha)$, where

$$h(\alpha) \equiv \frac{\|\alpha\|}{\sqrt{c}},$$

and $\|\alpha\|$ is the standard Euclidean norm.

For instance, suppose that α is distributed according to the (truncated) independent multivariate normal, where the density function is (for $\alpha \geq 0$)

$$f(\alpha) = \left(\frac{2}{\pi \sigma^2} \right)^{n/2} \exp\left\{ -\frac{\|\alpha\|^2}{2\sigma^2} \right\}.$$

(The coefficient in front of the exponential is chosen to make the function integrate to one on the positive orthant.) We can write this in the form (23) by writing

$$f_1(h) = \left(\frac{2}{\pi \sigma^2} \right)^{n/2} \exp\left\{ -\frac{ch^2}{2\sigma^2} \right\}, \qquad f_2(\alpha) \equiv 1.$$

[15] This is a slight simplification of the problem analyzed in Mirman and Sibley (1980). (They specified utility as $u(\alpha, x)$ and required that the marginal utility of each product i was increasing in the scalar α.)

[16] This result is implicit in the analysis of Mirman and Sibley (1980, Section 3), although it is not emphasized.

If we write

$$\Psi_n(h) = h - \frac{\int_h^\infty \hat{h}^{n-1} \exp\left\{-\frac{c\hat{h}^2}{2\sigma^2}\right\} d\hat{h}}{h^{n-1} \exp\left\{-\frac{ch^2}{2\sigma^2}\right\}},$$

then Proposition 3 states that the optimal demand function x^* is given by

$$\text{(29)} \qquad x^*(\alpha) \text{ maximizes: } \sum_{i=1}^{n} \alpha_i \sqrt{x_i} \text{ subject to } c \sum_{i=1}^{n} x_i \leq y^*(h(\alpha)),$$

where

$$\text{(30)} \qquad y^*(h) \underset{y \geq 0}{\text{maximizes}}: \Psi_n(h)\sqrt{y} - y.$$

(Since f_1 is log concave, both hazard rate conditions (21) and (22) are satisfied.) Expressions (29) and (30) imply

$$\text{(31)} \qquad x_i^*(\alpha) = \frac{\alpha_i^2 [\Psi_n(h(\alpha))]^2}{4c\|\alpha\|^2}$$

provided that $\Psi_n(h(\alpha)) \geq 0$, and that $x_i^*(\alpha) = 0$ otherwise.

Integrating the numerator of the fraction in Ψ_n by parts yields the difference equation (for $n > 2$):

$$\Psi_n(h) = h - \frac{(n-1)\sigma^2}{ch} + \frac{(n-2)\sigma^2}{ch^2} \Psi_{n-2}(h).$$

Starting values are $\Psi_2(h) = h - \sigma^2/(ch)$, and $\Psi_1(h)$ which has no closed-form expression. The form of the solution depends then on whether there is an even or an odd number of goods. In particular, when $n = 2$

$$\text{(32)} \qquad x_i^*(\alpha) = \frac{\alpha_i^2}{4c^2} \left[1 - \frac{\sigma^2}{\|\alpha\|^2}\right]^2$$

provided that $\|\alpha\| \geq \sigma$. (A consumer buys nothing if $\|\alpha\| < \sigma$.) A striking feature of this example is that it (almost always) involves *pure bundling*: if a consumer participates in the market at all she will buy both goods (except on the axes themselves). Since the efficient levels of outputs in this example are given by $x_i(\alpha) = \alpha_i^2/(4c^2)$, from (32) we see that consumers are served with suboptimal quantities, but that the optimal demand function is asymptotically efficient as $\alpha \to \infty$.

Similar solutions when utility takes this form may be found when tastes are distributed according to the multivariate Weibull distribution, i.e., when

$$f(\alpha)=k\left(\prod_{i=1}^{n}\alpha_i\right)\exp\left\{-\frac{\|\alpha\|^2}{2\sigma^2}\right\},$$

or when tastes follow the multivariate t distribution, i.e., when

$$f(\alpha)=k\left[1+\frac{\|\alpha\|^2}{d\sigma^2}\right]^{-(n+d)/2}.$$

5. CONCLUDING COMMENTS

This paper has partially analyzed the multiproduct nonlinear pricing problem. There were two main results. First we saw that in a wide variety of situations the firm found it optimal to exclude some low-demand consumers from the market (Proposition 1), a result which I believe to be a fairly general and distinctive feature of multidimensional mechanism design. Second, a method was found to solve the problem for a class of cases (Proposition 3). This involved following the procedure used in the single-product case after making use of a kind of "integration by parts" along rays from the origin. In effect the method was to solve the less constrained problem in which the necessary implementability constraints were ignored, and then to find conditions under which the solution satisfied all constraints. The main conditions that ensured this was valid were the two separability conditions (18) and (23).

In any future work it would be desirable to extend this analysis in three directions: First, it would be useful to know how to solve the problem when the separability conditions do not hold (see the discussion in Section 4.5 above). Second, it would be interesting to investigate how close optimal cost-based tariffs can get to the profits obtained from the optimal *general* tariff. Finally, it seems worthwhile to apply these techniques to other areas of mechanism design, for instance, to the optimal regulation of a multiproduct firm with unknown costs, or to the optimal design of multiproduct auctions.

Department of Economics, University of Southampton, Southampton SO17 1BJ, UK.

Manuscript received February, 1993; final revision received March, 1995.

APPENDIX

A. *Proof of Proposition 1*

To see this, suppose on the contrary that almost all consumers do have a positive surplus. Let $s^*(\alpha)$ be the optimal surplus function with associated tariff schedule $t^*(x)$ and demand function $x^*(\alpha)$. Define

$$A(\varepsilon)=\{\alpha\,|\,\alpha\in A \text{ and } s^*(\alpha)\le\varepsilon\}.$$

Then A being convex and s^* being convex, increasing, and continuous implies that each $A(\varepsilon)$ is compact and convex. If $t^*(\cdot)$ is optimal then $A(0)$ must be nonempty, for otherwise the firm would benefit by increasing the tariff at every point by some small fixed amount. The continuity of s^* implies that $A(\cdot)$ is a continuous family of sets, so that $A(\varepsilon) \to A(0)$ as $\varepsilon \to 0$.[17] If $V(A(\varepsilon))$ and $S(A(\varepsilon))$ denote respectively the volume and surface area of the set $A(\varepsilon)$, then

$$\text{(33)} \qquad \lim_{\varepsilon \to 0} : V(A(\varepsilon)) = V(A(0)); \qquad \lim_{\varepsilon \to 0} : S(A(\varepsilon)) = S(A(0)).$$

Now suppose that there are two distinct consumer types a and b contained in $A(0)$. Since s^* is convex and nonnegative, the point $\lambda a + (1-\lambda)b$, where $0 < \lambda < 1$, also receives zero surplus. Since A is assumed to be strictly convex, this point lies in the interior of A. Therefore, the set

$$\{\alpha \mid \alpha \in A \text{ and } \alpha \le \lambda a + (1-\lambda)b\}$$

has positive Lebesgue measure and so has positive measure with reference to the density function f. But since s^* is nondecreasing, this set is also contained in $A(0)$ and we reach the conclusion, a contradiction, that $A(0)$ has positive measure. We deduce that $A(0)$ contains one and only one point. Provided that $m \ge 2$ (in which case the volume and surface area of a single point is zero), expressions (33) imply

$$\text{(34)} \qquad \lim_{\varepsilon \to 0} : V(A(\varepsilon)) = \lim_{\varepsilon \to 0} : S(A(\varepsilon)) = 0.$$

We wish to show that the profit obtained from consumers in the set $A(\varepsilon)$, denoted by $\pi(\varepsilon)$, is small compared to ε. The profit obtained from such consumers is as given in (5) except that the domain of integration is $A(\varepsilon)$. Because s^* and $c(x^*(\alpha))$ are nonnegative, we obtain

$$\pi(\varepsilon) \le \int_{A(\varepsilon)} u(\alpha, x^*(\alpha)) f(\alpha)\, d\alpha.$$

Since u is homogeneous of degree one in α we have $u(\alpha, x^*(\alpha)) \equiv \alpha' u_\alpha(\alpha, x^*(\alpha))$, and since (4) implies that $\nabla s^* = u_\alpha$, we obtain

$$\text{(35)} \qquad \pi(\varepsilon) \le \int_{A(\varepsilon)} \alpha' \nabla s^*(\alpha) f(\alpha)\, d\alpha.$$

Next we use a standard theorem in multivariate calculus—the *Divergence Theorem*—which states that

$$\int_D \operatorname{div}(\mu)\, d\alpha = \int_{\partial D} \mu' \hat{n}\, dS,$$

where D is a closed convex set in R^m, ∂D is the surface of this set, $\hat{n}$ is the outward-pointing unit normal vector at a point on the surface of D, $\mu = (\mu_1, \dots, \mu_m)$ is an m-dimensional vector-valued function defined on D, $\operatorname{div}(\mu) \equiv \sum_{i=1}^m \partial \mu_i / \partial \alpha_i$ is the *divergence* of the vector field μ, and dS denotes integration over the surface ∂D. Applying the Divergence Theorem to the set $A(\varepsilon)$, setting

[17] The space of compact convex sets in R^m can be given a metric (the Hausdorff metric)—see Eggleston (1958). In particular, it makes sense to speak of a continuous family of such sets. The properties of this space needed in this proof are that if C is a compact convex set then the volume $V(C)$ and surface area $S(C)$ of such a set are well defined and both are continuous in C under the Hausdorff metric. The surface area of the set C is defined as a kind of derivative of the volume map:

$$S(C) = \lim_{\varepsilon \to 0} : \frac{V(C + \varepsilon B) - V(C)}{\varepsilon}$$

where B is the solid m-dimensional unit ball. In particular, the surface area of a single point is zero unless $m = 1$.

$\mu(\alpha) = \alpha s^*(\alpha) f(\alpha)$ and noting that

$$\operatorname{div}[\alpha s^*(\alpha) f(\alpha)] = \alpha' \nabla s^*(\alpha) f(\alpha) + s^*(\alpha) \operatorname{div}[\alpha f(\alpha)],$$

expression (35) becomes

$$(36) \qquad \pi(\varepsilon) \le \int_{\partial A(\varepsilon)} s^*(\alpha) f(\alpha) \alpha' \hat{n}\, dS - \int_{A(\varepsilon)} s^*(\alpha) \operatorname{div}[\alpha f(\alpha)]\, d\alpha.$$

Since f is assumed to be continuously differentiable on the compact set $A(\varepsilon)$, suppose that each of $f(\alpha)$, $f(\alpha)\alpha'\hat{n}$ and $\operatorname{div}[\alpha f(\alpha)]$ are bounded in $A(\varepsilon)$ for all sufficiently small ε in absolute value by B, say. Since $s^*(\alpha) \le \varepsilon$ in $A(\varepsilon)$, expression (36) implies

$$(37) \qquad \pi(\varepsilon) \le \varepsilon B[S(A(\varepsilon)) + V(A(\varepsilon))].$$

Finally, consider the change in profit caused by reducing the tariff t^* everywhere by $\varepsilon > 0$. This will cause all consumers in the set $A(\varepsilon)$ to exit, in which case profits bounded above by (37) are lost, but will cause all other consumers to pay ε more to the firm. Because there are no income effects in demand, this implies that the firm makes ε more profit from each of these consumers. Since the number of consumers who exit is no greater than $B \times V(A(\varepsilon))$, the total increase in profit as a result of the lump-sum reduction to t^* is at least

$$\Delta\pi \ge \varepsilon[1 - BV(A(\varepsilon))] - \varepsilon B[S(A(\varepsilon)) + V(A(\varepsilon))]$$

which has the same sign as

$$1 - B[2V(A(\varepsilon)) + S(A(\varepsilon))].$$

From (34) this implies that profits are strictly increased for all sufficiently small ε, contradicting the supposed optimality of t^*. *Q.E.D.*

B. Derivation of the Density of h

The following gives a geometric and relatively informal derivation—for a much more rigorous method of tackling such problems, see Muirhead (1982, Chapter 2).

The density function for h, $\phi(\cdot)$, is given by

$$\phi(t) = \lim_{\varepsilon \to 0} : \frac{1}{\varepsilon} \int_{\{\alpha | t \le h(\alpha) \le t + \varepsilon\}} f(\alpha)\, d\alpha,$$

which, given (23), is

$$(38) \qquad \phi(t) = f_1(t) \times \lim_{\varepsilon \to 0} : \frac{1}{\varepsilon} \int_{\{\alpha | t \le h(\alpha) \le t + \varepsilon\}} f_2(\alpha)\, d\alpha.$$

By changing variables $\alpha : \mapsto \alpha/t$ we obtain

$$(39) \qquad \int_{\{\alpha | t \le h(\alpha) \le t + \varepsilon\}} f_2(\alpha)\, d\alpha = t^m \int_{\{\alpha | 1 \le h(\alpha) \le 1 + \varepsilon/t\}} f_2(\alpha)\, d\alpha.$$

For small δ,

$$(40) \qquad \int_{\{\alpha | 1 \le h(\alpha) \le 1 + \delta\}} f_2(\alpha)\, d\alpha \approx \int_{\{\alpha | h(\alpha) = 1\}} f_2(\alpha) w(\alpha, \delta)\, dS$$

where $w(\alpha, \delta)$ is "width" of the region $\{\alpha | 1 \le h(\alpha) \le 1 + \delta\}$ at the point α, i.e., $w(\alpha, \delta)$ is the length of the straight line from α which is normal to the surface $h(\alpha) = 1$ and which just reaches the surface $h(\alpha) = 1 + \delta$. The right-hand integral in the above is the *surface* integral over the surface $h(\alpha) = 1$. Since the unit normal vector from the surface $h(\alpha) = 1$ at α is $\nabla h(\alpha)/\|\nabla h(\alpha)\|$, the width w is given by

$$h\left(\alpha+w\frac{\nabla h(\alpha)}{\|\nabla h(\alpha)\|}\right)=1+\delta,$$

which, by taking a first-order Taylor expansion of h, implies that

$$w(\alpha,\delta)\approx\frac{\delta}{\|\nabla h(\alpha)\|}.$$

Combining this with (40) gives

$$\int_{\{\alpha|1\le h(\alpha)\le 1+\delta\}} f_2(\alpha)\,d\alpha\approx\delta\int_{\{\alpha|h(\alpha)=1\}}\frac{f_2(\alpha)}{\|\nabla h(\alpha)\|}\,dS,$$

and so (39) implies that

$$\int_{\{\alpha|t\le h(\alpha)\le t+\varepsilon\}} f_2(\alpha)\,d\alpha\approx\varepsilon t^{m-1}\int_{\{\alpha|h(\alpha)=1\}}\frac{f_2(\alpha)}{\|\nabla h(\alpha)\|}\,dS.$$

Finally, from (40) we obtain the formula given in the text:

$$\phi(t)=kt^{m-1}f_1(t)$$

where

$$k=\int_{\{\alpha|h(\alpha)=1\}}\frac{f_2(\alpha)}{\|\nabla h(\alpha)\|}\,dS.$$

REFERENCES

ADAMS, W. J., AND J. L. YELLEN (1976): "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90, 475–498.

EGGLESTON, H. G. (1958): *Convexity*. Cambridge: Cambridge University Press.

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. Cambridge: MIT Press.

LAFFONT, J.-J., E. MASKIN, AND J.-C. ROCHET (1985): "Optimal Nonlinear Pricing with Two Dimensional Characteristics," in *Information, Incentives and Economic Mechanisms*, ed. by T. Groves, R. Radner, and S. Reiter. Oxford: Basil Blackwell.

MCAFEE, R. P., AND J. MCMILLAN (1988): "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 46, 335–354.

MCAFEE, R. P., J. MCMILLAN, AND M. WHINSTON (1989): "Multiproduct Monopoly, Commodity Bundling, and the Correlation of Values," *Quarterly Journal of Economics*, 103, 371–383.

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–196.

MIRMAN, L. J., AND D. S. SIBLEY (1980): "Optimal Nonlinear Prices for Multiproduct Monopolies," *Bell Journal of Economics*, 11, 659–670.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

——— (1976): "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, 6, 327–358.

——— (1986): "The Theory of Optimal Taxation," in *The Handbook of Mathematical Economics*, ed. by K. J. Arrow and M. D. Intriligator. Amsterdam: North-Holland.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons.

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

ROBERTS, K. (1979): "Welfare Considerations of Nonlinear Pricing," *Economic Journal*, 89, 66–83.

Rochet, J.-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics*, 16, 191–200.

Schmalensee, R. (1984): "Gaussian Demand and Commodity Bundling," *Journal of Business*, 57, S211–230.

Spence, A. M. (1980): "Multi-product Quantity-dependent Prices and Profitability Constraints," *Review of Economic Studies*, 47, 821–842.

Wilson, R. B. (1991): "Multiproduct Tariffs," *Journal of Regulatory Economics*, 3, 5–26 (and correction, page 211).

——— (1993): *Nonlinear Pricing*. Oxford: Oxford University Press.

——— (1995): "Nonlinear Pricing and Mechanism Design," in *The Handbook of Computational Economics*, Volume 1, ed. by H. Amman, D. Kendrick, and J. Rust. Elsevier Science Publishers, forthcoming.